METHOD FOR GENERATING AN IMPROVED IMAGE SIGNAL WHEN ESTIMATING THE MOTION OF IMAGE SEQUENCES, IN PARTICULAR A PREDICTION SIGNAL FOR VIDEO IMAGES USING MOTION-COMPENSATING PREDICTION

INS.   A1) ~~Background Information~~

The present invention relates to a method for generating an
improved image signal when estimating the motion of image

5   sequences, in particular a prediction signal for video
images using motion-compensating prediction, with motion
vectors, which, for each picture block of a current image,
indicate the position of the picture block used for the
prediction with respect to a chronologically preceding

10   reference image, being formed for picture blocks.

A2)
No. 0 558 922
European Patent ~~EP 0 558 922 B2~~ describes a method for
improving motion estimation in image sequences, in half-pel
accuracy, according to the full-search method. There, in a

15   first process step, the search area, and in a second process
step, the match block are filtered with the aid of an
additional digital filter which enables a raster shift of
the pixel raster by ¼ pel. Using this measure, a distortion
of the motion vector field can be ruled out.

20

A   In ~~the~~ "MPEG-4 Video Verification Model Version 7.0",
Bristol, April 1997, MPEG 97/N1642 in ISO/IEC JTC1/SC
29/WG11, an encoder and decoder for object-based coding of
video image sequences are specified. In this context, one no

25   longer encodes and transmits rectangular pictures of a fixed
size to the receiver, but instead, so-called "VIDEO OBJECTS"
(VO) of any shape and size. The image formation of such a VO
in the camera image plane at a specific instant is referred
to as a VIDEO OBJECTS PLANE (VOP). Consequently, the

relation between VO and VOP is equivalent to the relation between image sequence and image in the case of the transmission of rectangular pictures of fixed size.

5   The motion-compensating prediction in the verification model is carried out with the assistance of so-called "blockwise motion vectors" which, for each block of the size 8 x 8 or 16 x 16 pixels of the current image, specifies the position of the block used for the prediction in an already
10   transmitted reference image. In this context, the resolution of the motion vectors is limited to half of a pixel, pixels between the scanning raster (half-pixel position) being generated by a bilinear interpolation filtering from the pixels on the scanning raster (integer pixel position)
15   (Figure 1). In this case, + indicates the integer-pixel position, and O the half-pixel position. Interpolated values a, b, c, and d in half-pixel position are derived from the following relations:
a = A, b = (A + B)//2, c = (A + C)//2,
20   d = (A + B + C + D)//4, // indicating a rounded integer division.

~~Summary of the Invention~~

INS. A3>

25 A By applying the principles ~~measures~~ of the present invention, one can improve the quality of the prediction signal and, thus, the coding efficiency. In so doing, a greater local neighborhood is considered than in the case of bilinear interpolation, to generate pixels between the pixel scanning raster. The
30   aliasing-reducing interpolation filtering according to the present invention leads to an increased resolution of the motion vector and, consequently, to a prediction gain and an increased coding efficiency. In the present invention, the

2

FIR filter coefficients can be adapted to the signals to be coded, and be transmitted separately for each video object, thereby further increasing coding efficiency and enhancing the flexibility of the method.

5

In contrast to the design approach according to European Patent No. 0 558 922, there is no need to design any additional polyphase filter structures for intermediate positions having ¼ pel pixel resolution in the horizontal

10 and vertical directions.

By applying the principles of the present invention, the image sequence frequency of an MPEG-1 coder can be doubled from 25 Hz to 50 Hz, with the data rate remaining constant. In the

15 case of an MPEG-2 coder, the data rate can be reduced by up to 30%, with the image quality remaining constant.

Brief Description of the Drawings

Exemplary embodiments of the present invention are explained

20 in the following in greater detail, on the basis of the drawing, whose figures show:

Figure 2   an interpolation diagram for pixels between the
25   pixel scanning raster;
Figure 3   the structure of an FIR filter for interpolation;
Figure 4   a further interpolation having an even higher
           resolution.

30 Description of Exemplary Embodiments

In the case of the method according to the present invention, motion vectors are formed for picture blocks, the

3

motion vectors, for each picture block of a current image,
indicating the position of the picture block used for the
prediction with respect to a chronologically preceding
reference image.

5

The motion vectors for the prediction are determined in
three successive steps:
In a first search step, a motion vector is determined for
each picture block with pel accuracy in accordance with a

10   conventional method, for example, in accordance with the
full-search block matching method. In this context, the
minimum error criterion is determined for possible motion
positions, and the vector which best describes the motion of
the picture block is selected (European Patent, ~~EP 0 368 151~~ No. 0 368 151

15   ~~B1)~~ .

In a second search step, which, again, is based on such a
search for the minimum error criterion, an improved motion
vector is ascertained with sub-pel accuracy, starting out

20   from the motion vector ascertained in the first step, using
an aliasing-reducing interpolation filtering, with the aid
of a digital, symmetric FIR (finite impulse response)
filter. In the process, a higher resolution is selected than
in the first search step. Preferably, one selects a

25   resolution of a half pixel relative to the pixel raster.
Figure 2 illustrates the interpolation pattern for pixels b,
c, and d between the pixel raster, as derived from the
neighboring pixels A, B, C, D, E, F, G, H on the pixel
raster. + indicates the integer-pixel position, O the half-

30   pixel position, so that:

$$b = (CO1x(A_{-1} + A_{+1}) + CO2x (A_{-2} + A_{+2}) + CO3x (A_{-3} + A_{+3}) + CO4x (A_{-4} + A_{+4}))/256$$

$$c_i = (CO1x(A_i + E_i) + CO2x(B_i + F_i) + CO3x\ (C_i + G_i) + GO4x(D_i + H_i))/256$$

$$d = (CO1x(c_{-1} + c_{+1}) + CO2x(c_{-2} + c_{+2}) + CO3x\ (c_{-3} + c_{+3}) + CO4x(c_{-4} + c_{+4}))/256$$

The structure of the FIR interpolation filter used is apparent in Figure 3. Following each pixel position δp, it branches to a coefficient evaluator 1, 2, 3, etc., and has a summing device 10 at the output. As is apparent from the above relationships, a greater local neighborhood is considered when generating pixels between the scanning raster, than in the case of the bilinear interpolation according to the related art. In this context, the interpolation filter coefficients CO2, CO3, CO4 are determined in a way that minimizes the interpolation error performance. The coefficients can be determined directly using the known estimation method of the least root-mean-square error. From the minimization of the interpolation error performance, one obtains a linear system of equations, whose coefficients can be derived from the principle of orthogonality. A set of FIR filter coefficients optimized in this manner is given by the coefficients CO1 = 161/256, CO2 = -43/256, CO3 = 23/256, CO4 = -8/256.

In the third search step, starting from the motion vector determined with an accuracy of ½ pel, a local search is performed using a further interpolation filtering, taking the eight neighboring pixels as a basis, with resolution that is increased still further, preferably to ¼ pixel. As before, one selects the motion vector having the lowest prediction error performance.

Figure 4 shows the interpolation pattern for this. The

integer pixel positions are marked by X, the half-pixel positions by O, and the quarter-pixel positions by -. O indicates the best compensation with ½ pixel, and + the quarter-pixel search position.

The interpolation is carried out relative to the pixel raster, with a half-pixel resolution from the second search step, using filter coefficients CO1' = 1/2, CO2' = O, CO3' = O, CO4' = O.

The same previously introduced interpolation technique is used for the motion-compensating prediction.

If the processing is carried out within a coder having a reduced image format (SIF format within an MPEG1 coder or Q-CIF in an H.263 coder), but the original input format is used for the display, for example, CCIR 601[1] in the case of MPEG-1 or CIF in the case of H.263, a local interpolation filtering must be carried out as a post-processing. The described aliasing-compensating interpolation filtering can be used for this purpose as well.

To activate the aliasing-compensating interpolation using 1/4 resolution, activation bits can be inserted into an image-transmission bit stream.

To predict video objects, filter coefficients CO1 through CO4, and CO1' through CO4' can be separately conditioned for each of the video objects VO, and inserted into the image-transmission bit stream at the beginning of transmission of the video object in question.

For the encoding of a motion vector, the range of values of

6

the motion vector differences to be coded can be adapted to
the increased resolution.